干旱区地理
ARID LAND GEOGRAPHY

doi:10.12118/j.issn.1000－6060.2018.06.17

MLR 和 PLSR 的沙壤土盐分含量光谱检测对比研究[①]

王 涛[1,2]， 喻彩丽[1,2]， 姚 娜[1,2]， 王 斐[1,2]， 白铁成[1,3]

（1 塔里木大学信息工程学院，新疆 阿拉尔 843300； 2 新疆南疆农业信息化研究中心，新疆 阿拉尔 843300；
3 Gembloux Agro-Bio Tech，University of Liège，Gembloux 25030，Belgium）

摘 要： 为了快速有效检测南疆地区典型土壤（沙壤土）的盐分含量变化，利用光谱仪和电导仪测得南疆阿拉尔市红枣种植区盐渍土近红外高光谱和电导率数据，基于7种不同光谱预处理方法和2种特征波长选择算法，分别建立多元线性回归（MLR）和偏最小二乘回归（PLSR）的土壤盐分监测模型。结果表明：7种预处理方法中，归一化，多元散射，变量标准化和一阶导数能够有效提高土壤盐分的预测模型精度。基于多元逐步回归（SMR）波长选择方法的多元线性回归（SMLR）模型的 ${R_{val}}^2<0.9489$，$RPD<6.2949$，$RMSEP>0.4356$；基于连续投影算法（SPA）的多元线性回归（SPA-MLR）模型的 ${R_{val}}^2<0.9568$，$RPD<6.9221$，$RMSEP>0.3616$，预测结果要优于偏最小二乘回归（PLSR）模型，其中基于归一化处理后的 SMLR 和 SPA-MLR 的预测精度最为理想，分别为 ${R_{val}}^2=0.9792$，$RPD=9.9078$，$RMSEP=0.2876$ 和 ${R_{val}}^2=0.9805$，$RPD=10.50$，$RMSEP=0.2783$，而且筛选的特征波长较少。说明归一化是更有效的光谱预处理方法，多元线性回归（MLR）更适合建立南疆典型沙壤土盐分含量的预测模型。

关 键 词： 土壤电导率；多元线性回归；偏最小二乘回归方法；
中图分类号： S151.9 **文献标识码：** A **文章编号**

土壤盐渍化是新疆南部地区土地退化的主要问题，对南疆农业可持续发展以及生态环境的保护至关重要[1]。传统土壤盐渍化监测，依赖于土壤理化分析方法，不能快速有效的反映区域土壤盐渍化动态信息[2]，近年来通过土壤光谱信息估算土壤中全氮[3]、水分[4]、有机质[5]等研究较多，利用高光谱预测土壤盐渍化定量分析也逐步展开，遥感技术利用土壤的盐分光谱特征，能够快速、实时的获得盐渍化土壤的盐分信息[6-7]，因此，研究土壤盐分的光谱特征，建立基于光谱分析的土壤盐分定量反演模型是一项重要的工作。

土壤盐渍化的高光谱定量反演模型是土壤盐渍化高光谱遥感监测的重要研究内容之一。目前，已有学者利用不同的研究方法对不同地区的盐渍化土壤盐分进行了高光谱定量反演。常用的建模方法包括偏最小二乘回归、多元线性回归和主成分回归等。GOLDSHLEGE N[8]等通过获取番茄地高光谱反射率与土壤盐分含量，使用偏最小二乘法建立预测模型，取得了较好的预测精度，其中预测均方根误差 *RMSE* 是0.6%，决定系数 R^2 是0.87。NETO OR[9]等在巴西半干旱地区使用土壤电导率进行土壤盐渍化评价，对比使用主成分回归，多元线性回归和偏最小二乘法建立预测模型，其中偏最小二乘法获得了较好的预测精度，*RMSE* 是0.22，*RPD* 是2.21。彭杰等[10]通过分析南疆地区土壤的高光谱数据与盐分含量关系，基于连续统去除法所建模型的相对分析误差（*RPD*）达2.5以上。屈永华等[11]建立了内蒙古河套灌区土壤盐分与高光谱数据的偏最小二乘回归（PLSR）模型，验证数据的决定系数（R^2）为0.73，预测方差比为1.79。丁建丽等[12]通过分析新疆塔里木盆地北缘渭干河—库车河三角洲绿洲盐渍化土壤光谱信息与盐分关系，发现利用综合光谱指

① **收稿日期：** 2018－04－27； **修订日期：** 2018－07－28
基金项目： 国家自然科学基金项目（61501314，41561088）；塔里木大学校长基金项目（TDZKQN201614）
作者简介： 王涛（1982－）男，陕西西安人，塔里木大学讲师，研究生，主要从事遥感与数字农业技术研究．E-mail：wujiang0156@163.com
通讯作者： 白铁成（1982－）男，黑龙江绥化人，塔里木大学副教授，比利时列日大学博士研究生，主要从事遥感与数字农业技术研究．E-mail：baitiecheng1983@163.com

数来监测土壤盐渍化是非常有效的，预测的 *RMSE* 为0.009。段鹏程等[13]通过对新疆阜康500水库下游的盐渍化土壤实地定点取样和光谱测量，建立了土壤反射光谱与盐分含量之间的多元线性回归预测模型，在不同类型的土壤含盐量预测模型中，一阶导数建立的模型预测效果最优，R^2 均超过0.983。李亚莉等[14]以新疆典型盐渍化灌区为研究对象，获取不同质地的土壤水盐含量光谱曲线，采用偏最小二乘回归方法构建水盐反演模型，模型的决定系数 R^2 均大于0.65。

尽管已有研究利用土壤电导率和含盐量指标，取得了相对理想的反演效果，但不同区域的土壤结构、成分差异显著，不同类型土壤的高光谱响应和敏感波段也具有一定的差异性，不同方法建模的预测效果也有很大不同。因此，本文在总结前人研究基础上，以新疆阿拉尔沙壤土为研究区，依据高光谱数据与土壤电导率实测数据，对获取的土壤光谱数据进行多元散射校正、矢量归一化、变量标准化、移动平滑、SG平滑、一阶导和二阶导等不同预处理，分别使用多元逐步回归(SMR)与连续投影算法(SPA)筛选土壤电导率敏感特征波长，对比分析多元线性回归(MLR)和偏最小二乘回归法(PLSR)建立土壤电导率检测模型的精度和效率，旨在筛选和建立一种适合该区域的土壤盐分预测方法和模型，为利用遥感技术快速、无损监测新疆南部地区土壤盐渍化提供理论依据和实用方法。

1 材料与方法

1.1 数据采集

试验区位于新疆阿拉尔市，81°13′3″E，40°34′45″N，枣园土壤类型为沙壤土。分别采集土壤表层盐分已结晶的表层土和0～20 cm的正常土壤，按1∶99，2∶98，…，99∶1的比例混合共获得142份不同盐分的土壤样本，带回实验室自然风干，磨细过筛，用于光谱及理化分析测试用。因为土壤电导率已成为国际上广泛使用的盐分表示方式[15]，本研究使用电导率代表土壤盐分含量状况，电导率采用水土比为5∶1，利用DDS—307电导率仪直接进行测定。光谱采用Zolix Gaia Sorter近红外成像高光谱仪，光谱测定范围900～1 700 nm(实际测量到1 750 nm)，光谱分辨率5 nm，光谱采样点4 nm，共256个波段测定获取高光谱数据，每次测量前进行标准白板校正，每个样点重复采集10条光谱曲线，取平均值作为该样本的光谱反射率，剔除了2份异常样本后(光谱曲线缺失)，共获得140个有效样本，如表1是根据SPXY算法[16]挑选出的100份样品作为定标集，其他40份土壤样品作为预测集。

光谱曲线通过异常剔除和算数平均值处理后，获取140个样本的光谱吸收谱(图1)。

1.2 光谱数据预处理与建模方法

1.2.1 光谱数据预处理方法与特征波长选择方法

由于近红外光谱区的吸收强度弱，光谱信噪比低，谱带易重叠，为了消除仪器噪声等对光谱数据的影响以及准确分析土壤样品中的含盐量，有必要对土壤原始光谱进行预处理以提高检测的精度。对原始光谱数据分别进行了多元散射校正(MSC)、矢量归一化(VN)、变量标准化(SNV)、移动平滑(MA)、SG平滑(SG)一阶导(1－Der)和二阶导(2－Der)等预处理，其中MSC，SNV，VN可以校正样品因颗粒散射而引起的光谱误差，VN可以较好的改善样品稀疏时的预测结果，平滑处理可以有效平滑高频噪声，提高信噪比，一阶导数和二阶导数分别用于消除光谱中基线的平移和漂移，可有效消除其他背景干扰，分

表1 土壤校正集和预测集盐分含量统计

Tab.1 Statistics of soil electrical conductivity in calibration and validation set

数据集	样本数	最小值	最大值	平均值	标准差
校正集	100	0.272 0	10.540 0	5.656 1	2.813 1
预测集	40	1.144 0	9.870 0	1.144 0	2.542 2

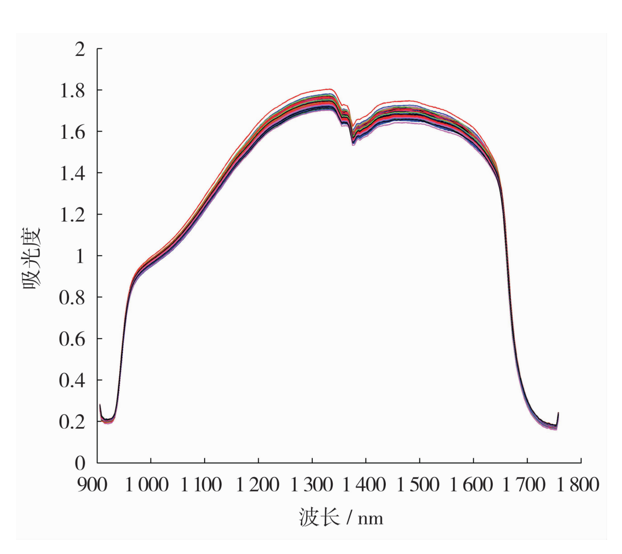


图1 土壤原始光谱曲线图

Fig.1 Original spectrum of soil

辨重叠峰,提高分辨率和灵敏度[17-18]。

土壤电导率敏感特征波长选择分别采用多元逐步回归(SMR)与连续投影算法(SPA)算法,其中SMR算法可集成在MLR建模算法中,构建多元线性回归模型(SMLR)。

1.2.2　建模方法及评价指标　SMLR是在多元线性回归的基础上发展起来的[15,19],逐步回归按照全局自变量对因变量的作用、显著程度或贡献率,逐个引入到回归方程中,剔除对因变量不显著的自变量[20-21],建立多元线性回归模型,对因变量进行预测。

PLSR吸取了主成分分析中提取主成分的思想,能够简化数据结构,主成分之间不但相关性最小,而且与因变量的相关性最大,同时,能够克服主成分分析对自变量有较强解释的能力,有效提取对系统解释性最强的综合变量,提高模型的估测能力。因而,本研究分别采用多元线性回归和偏最小二乘回归方法[22-23]构建土壤电导率反演模型,预测光谱特征值和土壤电导率之间的关系。

预测模型采用外部验证,对其稳定性和预测精度进行评价,采用建模决定系数(R_{cal}^2)、建模均方根误差(*RMSEE*)、预测决定系数(R_{val}^2)、预测均方根误差(*RMSEP*)、相对分析误差(*RPD*)为精度评价指标,R_{val}^2 和 *RPD* 越大,*RMSEP* 越小,说明预测效果越好[24]。

2　数据处理结果与分析

2.1　土壤电导率多元线性回归模型

对原始光谱分别进行VN、MSC、SNV、SG等预处理,在SMR特征波长选择的基础上,分别建立多元逐步线性回归(SMLR)模型,并进行验证。首先将样品电导率作为因变量,设定模型入选变量的 *F* 检验概率为0.05,模型剔除变量的 *F* 检验概率为0.1,将其对应的土壤原始光谱及其预处理数据作为自变量,建立SMLR模型,其预测结果如表2。基于7种预处理方法建模方程的 R_{cal}^2 和预测 R_{val}^2 都很高,而且都大于交叉验证的 R_{cv}^2。其中,归一化,多元散射校正,变量标准化的预测精度相对较好且优于原始光谱,预测的 *RMSEP* 在0.287 6和0.331 2之间,R_{val}^2 在0.969 0和0.979 2之间,*RPD* 在8.104 4和9.907 8之间,其中归一化处理后的预测精度最高(*RMSEP* = 0.287 6,R_{val}^2 = 0.979 2,*RPD* = 9.907 8)。

在原始光谱预处理基础上,使用SPA筛选的特征波长作为自变量,建立SPA-MLR多元线性回归模型,对应的土壤电导率作为因变量,预测结果如表3。对于7种预处理方法,建模方程的 R_{cal}^2 和预测 R_{val}^2 都很高,而且都大于交叉验证的 R_{cv}^2。移动平滑和二阶导数的预测精度较差,其它预处理的预测精度相对较好且优于原始光谱,模型的 *RMSEP* 在0.267 5和0.307 9之间,R_{val}^2 在0.968 9和0.980 5之间,*RPD* 在8.130 5和11.178之间,其中一阶导数预测精度最高(*RMSEP* = 0.267 5,R_{val}^2 = 0.980 5,*RPD* = 11.178)。尽管经过MSC,SNV,1 - Der和2 - Der预处理后SPA筛选的变量较多,但和SMLR的预测结果比较,SPA-MLR预测的 *RMSEP* 降低,R_{val}^2 和 *RPD* 提高,说明SPA-MLR的预测效果略好于SMLR方法。

2.2　土壤电导率偏最小二乘回归模型

原始光谱预处理基础上,直接使用PLSR建立

表2　基于逐步回归SMR的土壤电导率多元线性回归模型(SMLR)

Tab.2　Soil electrical conductivity multiple linear regression model based on SMLR

数据变换	特征波段个数	定标		交叉验证		预测		
		RMSEE	R_{cal}^2	*RMSECV*	R_{cv}^2	*RMSEP*	R_{val}^2	*RPD*
原始光谱	4	0.477 4	0.941 0	0.505 0	0.934 2	0.388 6	0.953 2	6.579 8
归一化	3	0.423 9	0.948 4	0.442 0	0.943 9	0.287 6	0.979 2	9.907 8
移动平滑	5	0.380 5	0.964 2	0.406 3	0.959 2	0.362 0	0.956 2	6.805 1
多元散射校正	3	0.415 5	0.952 8	0.435 0	0.948 3	0.331 2	0.969 0	8.104 4
变量标准化	3	0.397 0	0.956 9	0.416 6	0.952 6	0.331 4	0.969 4	8.154 9
卷积平滑	4	0.424 6	0.955 0	0.446 6	0.950 3	0.388 8	0.951 5	6.462 0
一阶导数	5	0.474 2	0.937 9	0.506 4	0.929 3	0.425 6	0.948 9	6.294 9
二阶导数	6	0.419 2	0.950 6	0.458 9	0.941 0	0.435 6	0.949 4	6.325 7

土壤电导率回归模型，预测效果如表 4。其中主成分数表示从土壤光谱曲线中提取的用于解释土壤电导率的成分个数，对于建模所需提取的主成分个数，可以通过交叉有效性检验来确定。基于 7 种预处理建模的 R_{val}^2 和预测 R_{cal}^2 都很高，而且都大于交叉验证的 R_{cv}^2。除了卷积平滑的预测精度相对较差，其他处理后的模型预测值都优于原始光谱，预测的 R_{val}^2 在 0.954 3 和 0.973 7 之间，*RMSEP* 在 0.322 8 和 0.400 4 之间，*RPD* 在 6.679 8 和 8.838 1 之间，说明在光谱不同预处理的模型预测精度存在较大差异。从预测的 R_{val}^2 来看归一化，移动平滑，多元散射校正，变量标准交化，一阶导数，二阶导数都大于 0.95，预测效果均较好，其中归一化处理后的精度最高（*RMSEP* = 0.322 8，R_{val}^2 = 0.973 7，*RPD* = 8.838 1）。

原始光谱预处理后，使用 SPA 算法筛选敏感波长建立偏最小二乘法回归模型（SPA-PLSR），SPA 筛选的特征波长作为自变量，对应的土壤电导率作为因变量，预测结果如表 5。对卷积平滑处理后的光谱建立 SPA-PLSR 模型，没有通过 0.05 的水平检

表 3　基于 SPA 选择特征波长的土壤电导率多元线性回归模型（SPA-MLR）

Tab. 3　Soil electrical conductivity multiple linear regression model based on SPA（SPA-MLR）

数据变换	特征波段个数	定标		交叉验证		预测		
		RMSEE	R_{cal}^2	*RMSECV*	R_{cv}^2	*RMSEP*	R_{val}^2	*RPD*
原始光谱	5	0.419 6	0.954 6	0.447 9	0.948 2	0.311 8	0.970 7	8.444 5
归一化	7	0.414 7	0.950 7	0.449 7	0.942 1	0.278 3	0.980 5	10.501 0
移动平滑	6	0.392 7	0.961 8	0.424 7	0.955 4	0.361 6	0.956 8	6.922 1
多元散射校正	14	0.375 0	0.961 6	0.437 3	0.948 1	0.286 3	0.977 0	9.791 1
变量标准化	14	0.372 2	0.962 2	0.430 6	0.949 5	0.304 9	0.974 1	9.152 4
卷积平滑	5	0.405 0	0.959 1	0.431 3	0.953 7	0.307 9	0.968 9	8.130 5
一阶导数	18	0.366 7	0.963 1	0.453 4	0.944 0	0.267 5	0.980 5	11.178 0
二阶导数	12	0.399 3	0.955 2	0.465 8	0.939 5	0.341 6	0.969 6	8.227 3

表 4　土壤电导率偏最小二乘回归模型（PLSR）

Tab. 4　Soil electrical conductivity partial least squares regression model（PLSR）

数据变换	定标		交叉验证		预测		主成份	
	RMSEE	R_{cal}^2	*RMSECV*	R_{cv}^2	*RMSEP*	R_{val}^2	*RPD*	
原始光谱	0.691 7	0.874 1	0.702 0	0.871 1	0.556 9	0.916 0	5.087 9	4
归一化	0.397 1	0.954 8	0.448 7	0.941 9	0.322 8	0.973 7	8.838 1	6
移动平滑	0.383 0	0.963 7	0.434 8	0.953 4	0.365 7	0.954 3	6.679 8	7
多元散射校正	0.408 7	0.954 3	0.437 0	0.947 8	0.344 7	0.966 7	7.852 6	4
变量标准化	0.406 5	0.954 8	0.434 5	0.948 4	0.343 5	0.967 0	7.877 5	4
卷积平滑	0.673 5	0.884 9	0.712 9	0.872 1	0.615 7	0.884 5	4.489 4	4
一阶导数	0.356 8	0.965 1	0.491 3	0.934 9	0.337 8	0.969 3	8.850 3	7
二阶导数	0.386 3	0.958 1	0.450 0	0.942 2	0.400 4	0.958 7	7.176 5	5

表 5　基于 SPA 选择特征波长的土壤电导率偏最小二乘回归模型（SPA-PLSR）

Tab. 5　Soil electrical conductivity partial least squares regression model based on SPA（SPA-PLSR）

数据变换	特征波段个数	定标		交叉验证		预测		主成份	
		RMSEE	R_{cal}^2	*RMSECV*	R_{cv}^2	*RMSEP*	R_{val}^2	*RPD*	
原始光谱	5	0.419 6	0.954 6	0.447 9	0.948 2	0.311 8	0.970 7	8.444 5	5
归一化	7	0.577 6	0.903 2	0.601 2	0.895 3	0.449 2	0.950 5	6.832 8	2
移动平滑	6	1.095 5	0.684 9	1.145 8	0.661 4	1.051 5	0.589 4	2.375 1	3
多元散射校正	14	0.412 0	0.953 6	0.463 3	0.941 4	0.355 6	0.963 9	7.793 3	5
变量标准化	14	0.422 1	0.951 3	0.448 7	0.945 0	0.370 9	0.960 8	7.206 0	4
一阶导数	18	0.424 9	0.950 3	0.454 4	0.943 2	0.324 2	0.970 1	8.563 1	4
二阶导数	12	0.472 1	0.937 1	0.494 6	0.930 7	0.446 8	0.947 7	6.309 3	4

验,因此没有建立预测模型。由表 5 可以看出,对于 6 种预处理方法,经过 SPA 特征波长选择后的建模预测精度都略低于原始光谱,尽管多元散射校正和一阶导数的预测精度相对较好,但是所选择的特征波长较多,计算量较大。这可能因为经过 SPA 选择后入选 PLSR 的主成分与直接使用 PLSR 的主成分相比较数目明显减少,丢失了部分有效信息导致模型的预测精度有所下降。

比较表 4 和表 5 发现在光谱预处理前提下进行 SPA 特征波长选择后建立的 PLSR 模型的预测精度要低于直接使用 PLSR 模型,但是基于原始光谱的 SPA-PLSR 模型的预测精度高于 PLSR 模型。分析原因经过 SPA 选择后,通过选择特征波长来代表全波段进行建模本身就会丢失一些有效信息,在通过 PLSR 建模筛选主成分后,有效信息又会丢失一些,导致模型的预测精度降低。比较表 3 和表 5 发现在光谱预处理前提下建立的 SPA-MLR 模型精度高于 PLSR,但基于原始光谱建立的模型预测精度却是一样的。分析原因 MLR 不会筛选主成分,有效信息相对 PLSR 丢失较少,因此精度相对较高。因为原始光谱存在噪声导致经过 SPA 筛选特征波长后与 PLSR 出现筛选一致的情况,2 种模型预测精度一致,从而也说明,光谱在没有经过预处理进行建模时,因为受到噪声的影响,建模算法失效,预测结果没有了可比性。

综上,基于归一化处理后使用多元逐步回归(SMLR)和基于连续投影算法的多元线性回归模型(SPA-MLR)的预测精度相对较好,并且筛选的特征波长也较少,如图 2 所示。

3　讨 论

不同算法的预测结果比较如表 6 所示,不同变换形式对土壤光谱信息有一定影响,进而影响估测模型的精度。

本文基于不同光谱预处理,采用 SPA 选择回归变量建立的 MLR 模型的预测精度要优于 SMR,其原因可能是 SPA 在选择特征波长的能力上要比 SMR 强,且 SPA 选择的特征波长要多于 SMR,多个相关波长能够较好的反应土壤有机质的光谱信息,从而可以较好估算土壤有机质含量,但是 SPA 选择的回归变量较多,计算比较耗时,SMR 建模仅有少量波长的光谱能进入模型,而且这些波长之间仍然存在一定的共线性,模型对于原始光谱信息利用较少。

直接使用 PLSR 方法建模的预测精度要优于 SPA-PLSR 模型。直接使用 PLSR,可入选的主成分个数要多于 SPA-PLSR,包含了更多的有效信息,分析原因是 PLSR 的基本作法是首先在自变量集中提出第一成分,然后建立因变量与第一成分的回归方程,如果回归方程已达到满意的精度,则算法中止。否则继续第二对成分的提取,直到能达到满意的精度为止。经过 SPA 选择特征波长后,入选的建模的变量有限,且在此基础上在进行PLSR回归,经过

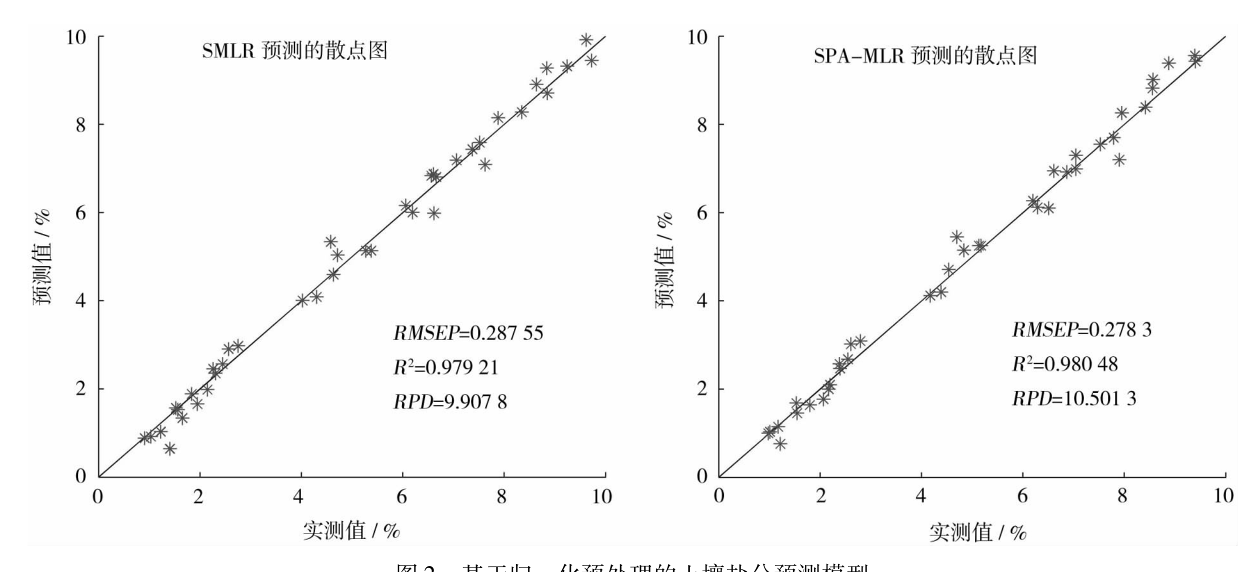


图 2　基于归一化预处理的土壤盐分预测模型

Fig. 2　The prediction model of soil salt content based on VN

表 6 不同建模算法预测精度比较

Tab. 6 Comparison of the prediction based on the different model

数据变换	SMLR 预测		SPA-MLR 预测		PLSR 预测		SPA-PLSR 预测	
	RMSEP	R_{val}^2	*RMSEP*	R_{val}^2	*RMSEP*	R_{val}^2	RMSEP	R_{val}^2
原始光谱	0.388 6	0.953 2	0.311 8	0.970 7	0.556 9	0.916 0	0.311 8	0.970 7
归一化	0.287 6	0.979 2	0.278 3	0.980 5	0.322 8	0.973 7	0.449 2	0.950 5
移动平滑	0.362 0	0.956 2	0.361 6	0.956 8	0.365 7	0.954 3	1.051 5	0.589 4
多元散射校正	0.331 2	0.969 0	0.286 3	0.977 0	0.344 7	0.966 7	0.355 6	0.963 9
变量标准化	0.331 4	0.969 4	0.304 9	0.974 1	0.343 5	0.967 0	0.370 9	0. 960 8
卷积平滑	0.388 8	0.951 5	0.307 9	0.968 9	0.615 7	0.884 5	/	/
一阶导数	0.425 6	0.948 9	0.267 5	0.980 5	0.337 8	0.969 3	0.324 2	0.970 1
二阶导数	0.435 6	0.949 4	0.341 6	0.969 6	0.400 4	0.958 7	0.446 8	0.947 7

0.05 的水平检验后,丢失了部分有效信息,入选的主成分个数也随之减少,导致模型的预测精度略低。

就 PLSR 与 SMLR 预测结果比较而言,除了导数预处理的情况,SMLR 建模精度要优于 PLSR 模型,这可能是由于逐步回归建模仅有少量波段的光谱吸收率能进入模型,而且逐步回归也可以减少变量间共线性问题,模型对于所选取的特征波长信息利用较多,逐步回归所选择的变量更能代表所有的波段。当然 PLSR 也能够处理自变量较多的情况且能降低变量之间的共线性,但在本文的实验结果中不及 SMLR 建模效果好。因此,SMLR 相对于 PLSR 可以更好地构建光谱与土壤盐分含量之间的关系模型。

4 结 论

利用高光谱成像技术获取南疆典型沙壤土光谱信息,并同步测定土壤电导率,通过 SMR 和 SPA 算法提取特征波段,结合 MLR 和 PLSR 建立回归模型,实现了对土壤电导率的快速检测。在文中所列的 7 种预处理方法中,较优的预处理方法是归一化,多元散射,变量标准化和一阶导数;预测模型 SPA-MLR 建模的预测精度要优于 SMLR,直接使用 PLSR 建模的预测精度要优于 SPA-PLSR,SMLR 和 SPA-MLR 模型的预测效果相对 PLSR 和 SPA-PLSR 较理想,表明相对于 PLSR 方法而言,MLR 更适合建立典型沙壤土电导率预测模型。

本文通过地面实测土壤光谱和电导率数据,综合分析对比了不同方法在新疆南疆阿拉尔地区沙壤土盐分检测时的精度和效率,建立的 SPA-MLR 预测模型的 $R_{val}^2 = 0.9805$, $RPD = 11.50$, $RMSEP = 0.2783$,为快速、准确地检测南疆枣园土壤电导率提供了一定的理论参考依据。此外,新疆南部地区土壤盐渍化严重,本研究在地面尺度准确获取的典型沙壤土的盐分敏感波长和监测模型,为后续使用卫星和无人机遥感进行大范围的有效监测提供了理论依据和实用模型,为构建多尺度、多类型土壤的盐渍化遥感监测提供基础数据和技术支持。但是不同类型土壤的含水量、质地、结构以及有机质含量等理化指标不同,光谱特征和建模方法也可能存在一定的差异,针对不同类型土壤的光谱预处理方法、特征波长选择方法和建模方法可在未来工作中加强对比研究。

参考文献(References)

[1] PANG G,TAO W,JIE L,et al. Quantitative model based on field-derived spectral characteristics to estimate soil salinity in Minqin County,China[J]. Soil Science Society of America Journal,2014,78(2):546 – 555.

[2] ZHAO Z,TASHPOLAT T,ZHANG F,et al. Spectral characteristics of soil salt content in typical oasis of Tarim River's middle reaches [J]. Journal of Natural Disasters,2012,21(5):72 – 78.

[3] 吴明珠,李小梅,沙晋明. 亚热带红壤全氮的高光谱响应和反演特征研究[J]. 光谱学与光谱分析,2013,33(11):3111 – 3115. [WU Mingzhu,LI Xiaomei,SHA Jinming. Spectral inversion models for prediction of red soil total nitrogen content in subtropical region (Fuzhou)[J]. Spectroscopy & Spectral Analysis,2013,33(11):3111 – 3115.]

[4] NAGY A,PETERRICZU,GALYA B,et al. Spectral estimation of soil water content in visible and near infra-red range[J]. 2014,3(3):163 – 171.

[5] ZHANG P,LI Y. Study on the comparisons of the establishment of two mathematical modeling methods for soil organic matter content based on spectral reflectance[J]. Spectroscopy & Spectral Analysis,2016,36(3):903 – 910.

[6] 张晓光,黄标,季峻峰,等. 基于可见近红外高光谱的东北盐渍土盐分定量模型研究[J]. 光谱学与光谱分析,2012,32(8):2075 – 2079. [ZHANG Xiaoguang, HUANG Biao, JI Junfeng, et al. Quantitative prediction of soil salinity content with visible-near

infrared hyper-spectra in northeast China[J]. Spectroscopy & Spectral Analysis,2012,32(8):2075 – 2079.]

[7] 杨爱霞,丁建丽,李艳红,等. 基于表观电导率与实测光谱的干旱区湿地土壤盐分监测[J]. 中国沙漠,2016,36(5):1365 – 1373.[YANG Aixia,DING Jianli,LI Yanhong,et al. Apparent electronic conductivity and measured spectral for monitoring soil salt content in Arid Lakeside Wetland[J]. Journal of Desert Research,2016,36(5):1365 – 1373.]

[8] GOLDSHLEGER N,CHUDNOVSKY A,BEN-BINYAMIN R. Predicting salinity in tomato using soil reflectance spectra[J]. International Journal of Remote Sensing,2013,34(17):6079 – 6093.

[9] NETO O R,TEIXEIRA A,LEAO R,et al. Hyperspectral remote sensing for detecting soil salinization using ProSpecTIR – VS aerial imagery and sensor simulation[J]. Remote Sensing,2017, 9(1):1 – 16.

[10] 彭杰,王家强,向红英,等. 土壤含盐量与电导率的高光谱反演精度对比研究[J]. 光谱学与光谱分析,2014,34(2):510 – 514.[PENG Jie,WANG Jiaqiang,XIANG Hongying,et al. Comparative study on hyperspectral inversion accuracy of soil salt content and electrical conductivity[J]. Spectroscopy and Spectral Analysis,2014,34(2):510 – 514.]

[11] 屈永华,段小亮,高鸿永,等. 内蒙古河套灌区土壤盐分光谱定量分析研究[J]. 光谱学与光谱分析,2009,29(5):1362 – 1366.[QU Yonghua,DUAN Xiaoliang,GAO Hongyong,et al. Quantitative retrieval of soil salinity using hyperspectral data in the region of Inner Mongolia hetao irrigation district[J]. Spectroscopy & Spectral Analysis,2009,29(5):1362 – 1366.]

[12] 丁建丽,伍漫春,刘海霞,等. 基于综合高光谱指数的区域土壤盐渍化监测研究[J]. 光谱学与光谱分析,2012,32(7):1918 – 1922.[DING Jianli,WU Manchun,LIU Haixia,et al. Study on the soil salinization monitoring based on synthetical hyperspectral index[J]. Spectroscopy & Spectral Analysis,2012,32(7):1918 – 1922.]

[13] 段鹏程,熊黑钢,李荣荣,等. 不同干扰程度的盐渍土与其光谱反射特征定量分析[J]. 光谱学与光谱分析,2017,37(2):571 – 576.[DUAN Pengcheng,XIONG Heigang,LI Rongrong,et al. A auantitative analysis of the reflectance of the saline soil under different disturbance extent[J]. Spectroscopy & Spectral Analysis,2017,37(2):571 – 576.]

[14] 李亚莉,乔江飞,董天宇,等. 不同质地盐渍化土壤水盐含量的高光谱反演[J]. 应用生态学报,2016,27(12):3807 – 3815.[LI Yali,QIAO Jiangfei,DONG Tianyu,et al. Hyperspectral inversion of soil water and salt content in soils with different textures[J]. Chinese Journal of Applied Ecology,2016,27(12):3807 – 3815.]

[15] 刘娅,潘贤章,王昌昆,等. 基于可见 – 近红外光谱的滨海盐土土壤盐分预测方法[J]. 土壤学报,2012,49(4):824 – 829.[LIU Ya,PAN Xianzhang,WANG Changkun,et al. Prediction of coastal saline soil salinity based on vis-nir reflectance spectroscopy[J]. Acta Pedologica Sinica,2012,49(4):824 – 829.]

[16] ZHAN X Y,ZHAO N,LIN Z Z,et al. Effect of algorithms for calibration set selection on quantitatively determining asiaticoside content in centella total glucosides by near infrared spectroscopy[J]. Spectroscopy & Spectral Analysis,2014,34(12):3267 – 3272.

[17] 尼珍,胡昌勤,冯芳. 近红外光谱分析中光谱预处理方法的作用及其发展[J]. 药物分析杂志,2008,28(5):824 – 829.[NI Zhen,HU Changqin,FENG Fang. Progress and effect of spectral data pretreatment in NIR analytical technique[J]. Chinese Journal of Pharmaceutical Analysis,2008,28(5):824 – 829.]

[18] 夏俊芳,李培武,李小昱,等. 不同预处理对近红外光谱检测脐橙 VC 含量的影响[J]. 农业机械学报,2007,38(6):107 – 111.[XIA Junfang,LI Peiwu,LI Xiaoyu,et al. Effect of different pretreatment method of nondestructive measure vitamin C content of umbilical orange with near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery,2007,38(6):107 – 111.]

[19] 刘娅,潘贤章,王昌昆,等. 土壤湿润条件下基于光谱对称度的盐渍土盐分含量预测[J]. 光谱学与光谱分析,2013,33(10):2771 – 2776.[LIU Ya,PAN Xianzhang,WANG Changkun,et al. Predicting soil salinity based on spectral symmetry under wet soil condition[J]. Spectroscopy & Spectral Analysis,2013,33(10):2771 – 2776.]

[20] 陈文倩,丁建丽,谭娇,等. 干旱区绿洲植被高光谱与浅层土壤含水量拟合研究[J/OL]. 农业机械学报,2017,(9):1 – 16 [CHEN Wenqian,DING Jianli,TAN Jiao,et al. Analysis vegetation of hyperspectral reflectance and shallow soil water content in arid area[J]. Transactions of the Chinese Society for Agricultural Machinery,2017,(9):1 – 16.]

[21] 侯艳军,塔西甫拉提·特依拜,买买提·沙吾提,等. 荒漠土壤有机质含量高光谱估算模型[J]. 农业工程学报,2014,30(16): 113 – 120.[HOU Yanjun,TIYIP Tashpolat,SAWUT Mamat,et al. Estimation model of desert soil organic matter content using hyperspectral data[J]. Transactions of the Chinese Society of Agricultural Engineering,2014,30(16):113 – 120.]

[22] 彭翔,胡丹,曾文治,等. 基于 EPO – PLS 回归模型的盐渍化土壤含水率高光谱反演[J]. 农业工程学报,2016,32(11):167 – 173.[PENG Xiang,HU Dan,ZENG Wenzhi,et al. Estimating soil moisture from hyperspectra in saline soil based on EPO-PLS regression[J]. Transactions of the Chinese Society of Agricultural Engineering,2016,32(11):167 – 173.]

[23] ZENG W Z,HUANG J S,XU C,et al. Hyperspectral reflectance models for soil salt content by filtering methods and waveband selection[J]. Ecological Chemistry & Engineering S,2016,23(1):117 – 130.

[24] FARIFTEH J,MEER F V D,ATZBERGER C,et al. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN)[J]. Remote Sensing of Environment,2007,110(1):59 – 78.

A comparison of the salt content in sandy soil between the MLR model and PLSR model

WANG Tao[1,2], YU Cai-li[1,2], YAO Na[1,2], WANG Fei[1,2], BAI Tie-cheng[1,3※]

(1 *College of Information Engineering*, *Tarim University*, *Aral* 843300, *Xinjiang*, *China*;

2 *South Xinjiang Agricultural Informatization Reaserch Center*, *Xinjiang Aral* 843300, *Xinjiang*, *China*;

3 *Gembloux Agro-Bio Tech*, *University of Liège*, *Gembloux* 25030, *Belgium*)

Abstract: In order to monitor typical soil salt content (sandy loam soil) in South Xinjiang, China quickly and effectively, and to improve the precision of the soil salt content estimation model by removing the noise of soil hyperspectral absorbance, this paper focused on the inversion relationship between soil spectrum and electrical conductivity (EC) by using multiple spectral pretreatment methods, and then the multiple linear regression (MLR) and the partial least squares regression (PLSR) modelling were applied to establish the salt content model based on the hyperspectral analysis technique. The effective and predictive capacities of different models were validated. This study took the typical arid area in South Xinjiang as the research object, obtained the hyperspectral data and EC by using Near-infrared spectrometer (Zolix Gaia Sorter) and conductivity meter (DDS – 307), 142 soil samples at 0 ~ 20 cm depth were collected and these samples were highly representative for the EC. Seven pretreatment methods were used to pretreat the original spectral data, such as vector normalization (VN), multiplicative scatter correction (MSC), standard normal variate (SNV), moving-average (MA) smoothing, Savitzky-Golay (SG) smoothing, first derivative (1 – Der) and second derivative (2 – Der), then the characteristic wavelengths were extracted with stepwise multiple regression (SMR) and successive projection algorithm (SPA), which were used as input variables of MLR and PLSR modeling. The results showed that the optimum pretreatment methods were VN, MSC, SNV and 1 – Der. According to different pretreatments, in the stepwise multiple linear regression (SMLR) prediction model ${R_{val}}^2$ *was greater than* 0.95, *RPD* was greater than 6.2, and *RMSEP* was less than 0.44. In the multiple linear regression model based on the successive projections algorithm (SPA-MLR) ${R_{val}}^2$ was greater than 0.96, *RPD* was greater than 6.9, and *RMSEP* was less than 0.36, which were better than those in SMLR. In the partial least squares regression prediction model ${R_{val}}^2$ was greater than 0.88, *RPD* was greater than 4.4, *RMSEP* was less than 0.62 and in the partial least squares regression model based on the successive projections algorithm (SPA-PLSR) ${R_{val}}^2$ was greater than 0.59, *RPD* was greater than 2.4, and *RMSEP* was less than 1.1 which were less than those in SMLR and SPA-MLR. The best prediction of SMLR and SPA-MLR after vector normalization (VN) was as the follows: *RMSEP* = 0.287 6, ${R_{val}}^2$ = 0.979 2, *RPD* = 9.907 8 and *RMSEP* = 0.278 3, ${R_{val}}^2$ = 0.980 5, *RPD* = 11.50 which had fewer characteristic wavelengths. So the VN is a best effective pretreatment method. It is more suitable to establish the prediction model of soil conductivity in jujube tree by MLR with PLSR. It could be an important part in future researches how to choose the right algorithm to analyze the soil salt content with the spectral data.

Key words: soil electrical conductivity; multiple linear regression; Partial least squares regression; Hyperspectral